

Viral Bioinformatics Resource Center

www.biovirus.org

Elliot Lefkowitz

UAB

Chris Upton

UVic

3rd BRC meeting

Virus Family Databases

	<i>Virus Family</i>	<i>Genomes</i>	<i>Genes</i>	<i>Genera</i>	<i>Species</i>	<i>Strains</i>
VBRC Databases	Arenaviridae	57	182	1	31	47
	Bunyaviridae	118	147	4	42	107
	Filoviridae	13	102	1	4	13
	Flaviviridae	269	2713	1	29	269
	Paramyxoviridae	139	1079	8	47	139
	Poxviridae	52	10216	10	27	52
	Togaviridae	52	583	1	22	52
		700	15022	26	202	679

Topics:

- a) **Genome Annotation Transfer Utility (GATU):** Rapid annotation of viral genomes using a closely related reference genome.
- b) **Longest Common Strings (LCS):** Finding common sequences in unaligned genomes.
- c) **Annotation of nucleotide ambiguities in viral meta-genomes:**
A question!

Viral Bioinformatics Resource Center

Welcome

The Viral Bioinformatics Resource Center (VBRC) at the University of Alabama at Birmingham and the University of Victoria is one of eight NIH/NIAID-funded Bioinformatics Resource Centers nationwide providing comprehensive web-based genomics resources to the scientific community conducting basic and applied research on microorganisms selected from the NIH/NIAID Category A, B, and C priority pathogens that are regarded as possible bioterrorist threats or as emerging or re-emerging infectious diseases. The VBRC is concerned primarily with viruses encompassing the families *Arenaviridae*, *Bunyaviridae*, *Flaviviridae*, *Filoviridae*, *Paramyxoviridae*, *Poxviridae*, and *Togaviridae*.

If you are interested in	Go here first
Web-based access to data and analysis tools	The VBRC Analysis Web
In-depth bioinformatics analytical tools	The VBRC Analytical Workbench

For information on all NIAID's Bioinformatics Resource Centers, please visit [BRC-Central](#).



Last Modified: 01/13/2006 09:03:14

[teams & people](#) | [about the vbrc](#) | [documents](#) | [feedback](#)

Copyright © 2004-2006 VBRC Consortium. All rights reserved. [Privacy Statement](#)



Google Search

- Search WWW
- Search biovirus.org
- Search athena.bioc.uvic.ca



Workbench

[VOCS](#)
[VGO](#)
[Base By Base](#)
[JDotter](#)
[ReHAB](#)
[GATU](#)
[SFS](#)
[GFS](#)
[NAP](#)
[GraphDNA](#)
[Hydrophobicity](#)
[Codon Statistics](#)

Web Tools

[Laj Gallery](#)
[BLAST](#)
[EMBOSS](#)

Information

[About VBCa](#)
[Virus Families](#)
[Workbench](#)
[Requirements](#)
[Report Bugs](#)
[Request Features](#)
[DB of Virologists \(DoV\)](#)
[Site Statistics](#)

Welcome

This resource

Provides access to viral genomes and a variety of tools for comparative genomic analyses. At the heart of the system is VOCs (Virus Orthologous Clusters), a database with built-in tools that allows users to retrieve and analyze the genes, gene families, and genomes of different virus families. The database is the source of information for other programs of the workbench for whole genome alignments, genome display, or gene/protein sequence analysis. Many of these tools can also be used with user-provided sequence data. The workbench tools are Java-based and user-friendly to allow all users, regardless of computer skill-level, to access and analyze the data.

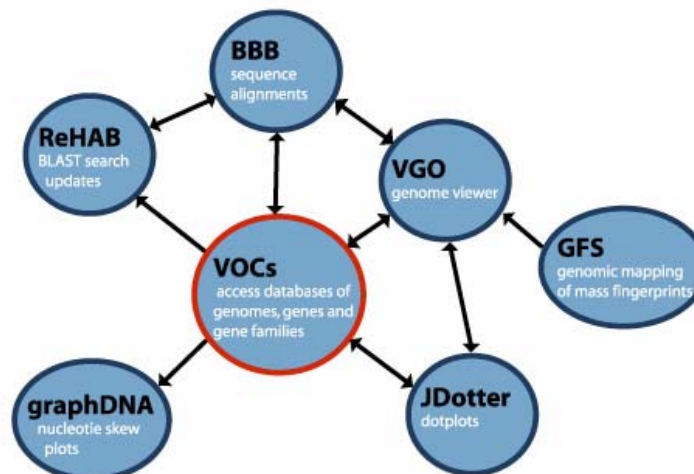
To start

There are two ways to start an application from the workbench:

- Select an application, and then choose its associated database when the application starts.
- Select a database, and then start an application using the afore-selected database.

[Quick Start Tutorial](#)

The figure below shows the interconnectivity between the workbench tools and databases at VBCa



Databases

dsDNA viruses

[Adenoviridae](#)
[Asfarviridae](#)
[Baculoviridae](#)
[Herpesviridae](#)
[Iridoviridae](#)
[Poxviridae](#)

+ ssRNA viruses

[Coronaviridae](#)
[Flaviviridae](#)
[Togaviridae](#)

- ssRNA viruses

[Arenaviridae](#)
[Bunyaviridae](#)
[Filoviridae](#)
[Paramyxoviridae](#)

Bacterial plasmids

[Plasmids](#)

Note:

Subscribe to keep informed of the latest developments at this website or enter your information in DoV. please click [Members Login](#) at DB of Virologists (DoV) site to create yourself an account.

[Home](#)[VBRC](#)[News](#)[Forum](#)[Links](#)[Contact](#)

Workbench

[VOCS](#)
[VGO](#)
[Base By Base](#)
[JDotter](#)
[ReHAB](#)
[GATU](#)
[SFS](#)
[GFS](#)
[NAP](#)
[GraphDNA](#)
[Hydrophobicity](#)
[Codon Statistics](#)

Web Tools

[Laj Gallery](#)
[BLAST](#)
[EMBOSS](#)

Information

[About VBCa](#)
[Virus Families](#)
[Workbench](#)
[Requirements](#)
[Report Bugs](#)
[Request Features](#)
[DB of Virologists \(DoV\)](#)
[Site Statistics](#)

Genome Annotation Transfer Utility



A warning message appears when running this software for the first time, since these programs access your computer. There is no cause for concern; we recommend that you simply click the start button. For more information please read: [Warnings Dialog](#)

[Launch Program](#)

Genome Annotator annotates a genome based on a very closely related reference genome. The proteins/mature peptides of the reference genome are BLASTed against the genome to be annotated in order to find the genes/mature peptides in the genome to be annotated.

Documentation

[Manual & Tutorial](#)
[Click here...](#)

The goal is to annotate the majority of the new genome's genes in a single step.

References ...

If your system does not launch Genome Annotation Transfer Utility, please see [Java Web Start Setup](#) and [Java Web start Tips](#) for help.

[license](#)[teams & people](#)[about the vbrc](#)[feedback](#)

This page has been visited 36 times since October 27, 2005

Databases

dsDNA viruses

[Adenoviridae](#)
[Asfarviridae](#)
[Baculoviridae](#)
[Herpesviridae](#)
[Iridoviridae](#)
[Poxviridae](#)

+ ssRNA viruses

[Coronaviridae](#)
[Flaviviridae](#)
[Togaviridae](#)

- ssRNA viruses

[Arenaviridae](#)
[Bunyaviridae](#)
[Filoviridae](#)
[Paramyxoviridae](#)

Bacterial plasmids

[Plasmids](#)

Note:

Subscribe to keep informed of the latest developments at this website or enter your information in DoV, please click [Members Login](#) at DB of Virologists (DoV) site to create yourself an account.

Genome Annotation Transfer Utility

Genome Annotation Transfer Utility

A warning message appears when running this software for the first time, since these programs access your computer. There is no cause for concern; we recommend that you simply click the start button. For more information please read: [Warnings Dialog](#)

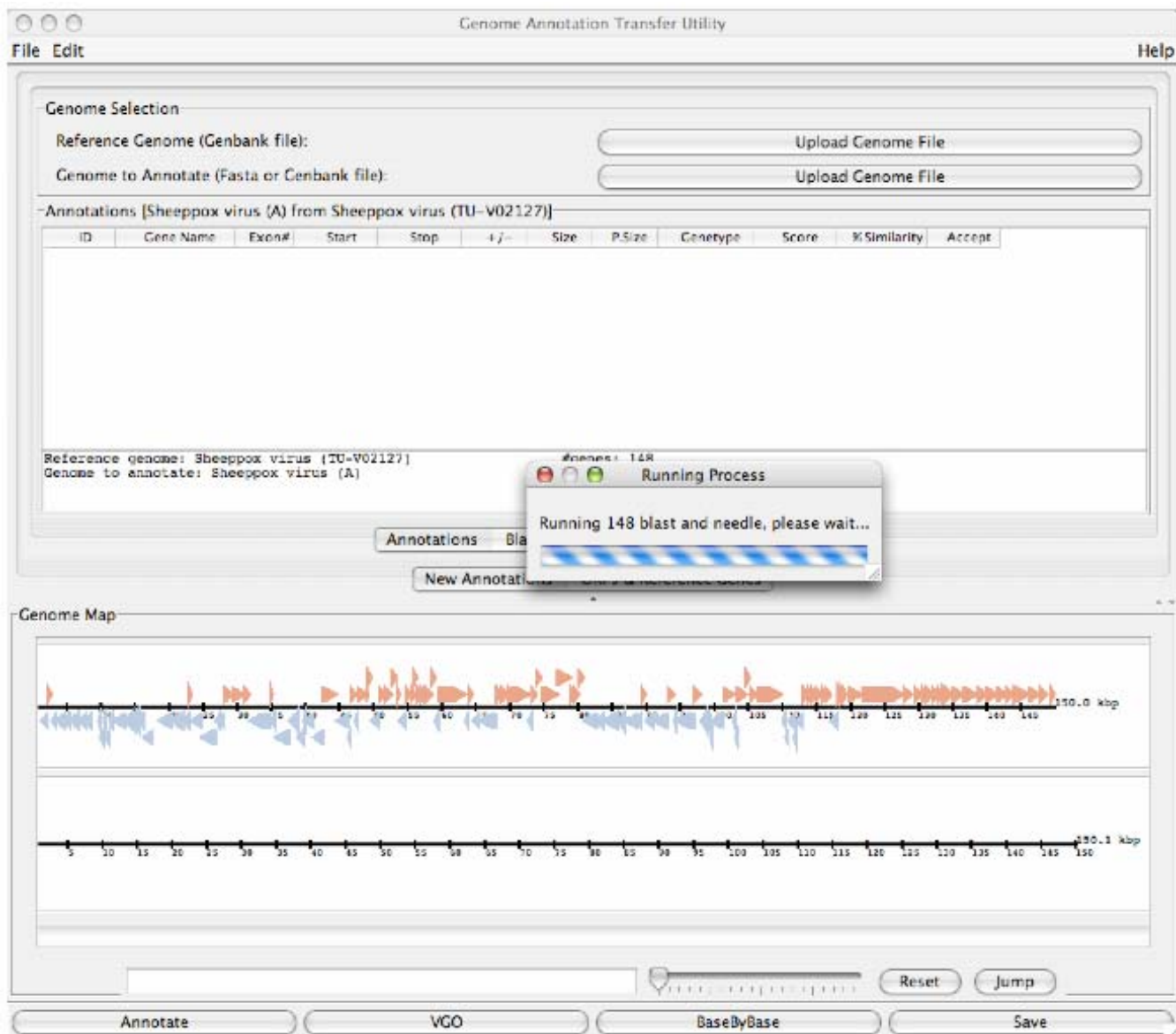
[Launch Program](#)

Genome Annotator annotates a genome based on a very closely related reference genome. The proteins/mature peptides of the reference genome are BLASTed against the genome to be annotated in order to find the genes/mature peptides in the genome to be annotated.

Documentation
[Manual & Tutorial](#)
[Click here...](#)

The goal is to annotate the majority of the new genome's genes in a single step.

GATU writes GenBank files.





Genome Selection

Reference Genome (Genbank file):

Upload Genome File

Genome to Annotate (Fasta or Genbank file):

Upload Genome File

Annotations [Sheeppox virus (A) from Sheeppox virus (TU-V02127)]

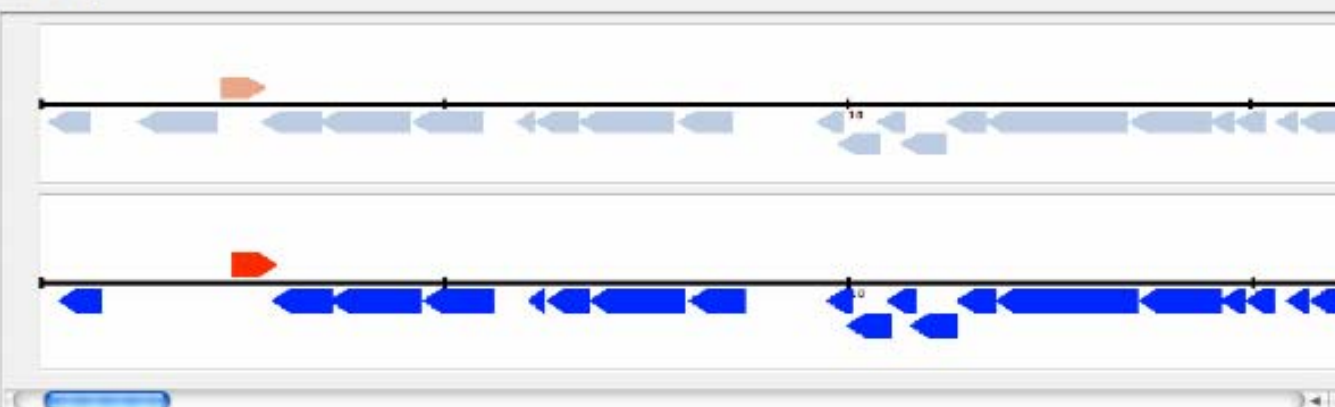
ID	Gene Name	Exon#	Start	Stop	+/-	Size	PSize	Genetype	Score	% Simila...	Accept
1 01		0	243	722	-	480	480	GENE	277.3	100	<input checked="" type="checkbox"/>
2 02		0	1367	2089	-	723	936	GENE	466.6	77.2	<input type="checkbox"/>
3 03		0	2372	2878	+	507	507	GENE	304.5	100	<input checked="" type="checkbox"/>
4 04		0	2889	3584	-	696	696	GENE	434.1	100	<input checked="" type="checkbox"/>
5 05		0	3607	4674	-	1068	1068	GENE	660.3	100	<input checked="" type="checkbox"/>
6 06		0	4756	5583	-	828	828	GENE	505.9	99.6	<input checked="" type="checkbox"/>
7 07		0	6056	6211	-	156	156	GENE	96.9	100	<input checked="" type="checkbox"/>
8 08		0	6284	6772	-	489	489	GENE	310.4	99.4	<input checked="" type="checkbox"/>

atgtcatcattaaacttttttctatttttctgttttcaactatctatttttactgtactgttaagcgggaatatcaataaagcgtgcacagagaaagaaatatacatggaatcgaagtag
M S B L T P F I P C L Q L F I F T S C V S G I S I K R C T E E E N N T W E I E V Q

Annotations Blast Alignment(s) Needle Alignment

New Annotations ORFs & Reference Genes

Genome Map



Ref: SPPV_02 [1/1] complement(1232..2167)

1232 2167

Reset

Jump

Annotate

VGO

BaseByBase

Save

```

Genome Annotation Trans
File Edit

#=====
#
# Aligned sequences: 2
# 1: SPPV_02
# 2: 02
# Matrix: /ca.virology/services/needle/matrices/EBLOSUM65
# Gap_penalty: 10.0
# Extend_penalty: 0.05
#
# Length: 311
# Identity: 240/311 (77.2%)
# Similarity: 240/311 (77.2%)
# Gaps: 71/311 (22.8%)
# Score: 1283.0
#
#=====
SPPV_02      1  MITSSTKNLFFFVNINSLDKRKRSFSFLSYVKILERKKRFVFKFLVNNRK      50
02           1                                           0
SPPV_02     51  KSSVSRWFVSTIRHKKQPTNMSSLTFFIFCLQLFIFTSTVSGISIKRCT      100
02           1                               MSSLTFFIFCLQLFIFTSTVSGISIKRCT      29
SPPV_02     101 EEENNTWEIEVGLCIQTENFRAIKTGCYKIQGPGLLTEGNGFKIFAHDD      150
02           30 EEENNTWEIEVGLCIQTENFRAIKTGCYKIQGPGLLTEGNGFKIFAHDD      79
SPPV_02     151 CSKEKTQNNFILDVNEAVYALGKYVYMEISTSNITTLNSLPQCAKRISL      200
02           80 CSKEKTQNNFILDVNEAVYALGKYVYMEISTSNITTLNSLPQCAKRISL      129
SPPV_02     201 SISCDQVTTEMKSYVESVSFKDYDLEFVITTDISCVKHVSSSVIVRNECE      250
02           130 SISCDQVTTEMKSYVESVSFKDYDLEFVITTDISCVKHVSSSVIVRNECE      179
SPPV_02     251 KKYISTGKKIFGFNNKIDCSAVKFSEHVNYLKTCSVGKFD RKKYYEHQHN      300
02           180 KKYISTGKKIFGFNNKIDCSAVKFSEHVNYLKTCSVGKFD RKKYYEHQHN      229
SPPV_02     301 YIKKIFHHNEL      311
02           230 YIKKIFHHNEL      240
#=====
```



File Edit

```
Query = 001.00;SPPV_02;0;0;;Sheeppox virus (TU-V02127)
      (311 letters)

>0;SPPV-A;0;SPPV-A;Sheeppox virus strain A
      Length = 150057

Score = 516.924 bits (1330), Expect = 3.38273e-149
Identities = 268/312 (85.9%), Positives = 277/312 (88.8%), Gaps = 4/312 (1.0%)
Frame = -3

Query: 1   MITSSTKNLFFFVNINSLDKRKRSFSFLSYVKILERKKRFVFK-FLVNNRKKSSVVSRLF 59
          MITSSTKNLFF      L ++K+ F F      K      +K+  F+ F      +KK S +
Sbjct: 1370 MITSSTKNLFFC*Y--KLLRQKKKFQFSIICKNP*EEKKVCFQVFS***KKKFSSIEVVC 1547

Query: 60  VSTIRHKKQPTNMSSLTFFIFCLQLFIFTSTVSGISIKRCTEEENNTWEIEVGLCIQTEN 119
          V      +KQPTNMSSLTFFIFCLQLFIFTSTVSGISIKRCTEEENNTWEIEVGLCIQTEN
Sbjct: 1548 VDNTS-QKQPTNMSSLTFFIFCLQLFIFTSTVSGISIKRCTEEENNTWEIEVGLCIQTEN 1726

Query: 120  FRAIKTGCIQGP GGLLTEGNGFKIFAHDDCSKEKTQNNFILDVNEAVYALGKYVYME 179
          FRAIKTGCIQGP GGLLTEGNGFKIFAHDDCSKEKTQNNFILDVNEAVYALGKYVYME
Sbjct: 1727 FRAIKTGCIQGP GGLLTEGNGFKIFAHDDCSKEKTQNNFILDVNEAVYALGKYVYME 1906

Query: 180  ISTSNITTLSLNPQCAKRISLSISCDQVTTEMKSYVESVSFKDYDLEFVITTDISCVKHV 239
          ISTSNITTLSLNPQCAKRISLSISCDQVTTEMKSYVESVSFKDYDLEFVITTDISCVKHV
Sbjct: 1907 ISTSNITTLSLNPQCAKRISLSISCDQVTTEMKSYVESVSFKDYDLEFVITTDISCVKHV 2086

Query: 240  SSSVIVRNECEKKYISTGKKIFGFNNKIDCSAVKFSEHVNYLKTCSVGKFDRKKYYEHQH 299
          SSSVIVRNECEKKYISTGKKIFGFNNKIDCSAVKFSEHVNYLKTCSVGKFDRKKYYEHQH
Sbjct: 2087 SSSVIVRNECEKKYISTGKKIFGFNNKIDCSAVKFSEHVNYLKTCSVGKFDRKKYYEHQH 2266

Query: 300  NYIKKIFHHNEL 311
          NYIKKIFHHNEL
Sbjct: 2267 NYIKKIFHHNEL 2296
```



Genome Selection

Reference Genome (Genbank file):

Upload Genome File

Genome to Annotate (Fasta or Genbank file):

Upload Genome File

Annotations [Sheeppox virus (A) from Sheeppox virus (TU-V02127)]

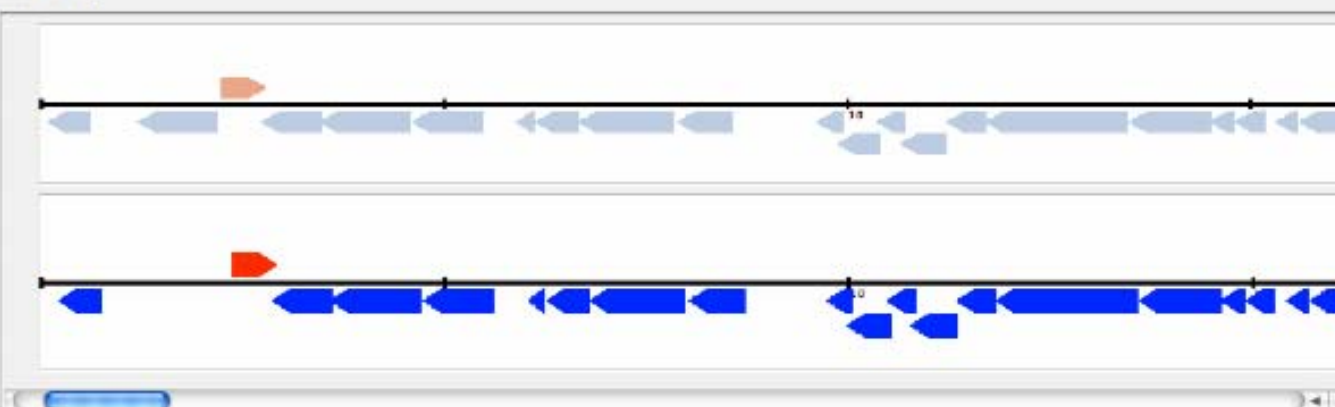
ID	Gene Name	Exon#	Start	Stop	+/-	Size	PSize	Genetype	Score	% Simila...	Accept
1 01		0	243	722	-	480	480	GENE	277.3	100	<input checked="" type="checkbox"/>
2 02		0	1367	2089	-	723	936	GENE	466.6	77.2	<input type="checkbox"/>
3 03		0	2372	2878	+	507	507	GENE	304.5	100	<input checked="" type="checkbox"/>
4 04		0	2889	3584	-	696	696	GENE	434.1	100	<input checked="" type="checkbox"/>
5 05		0	3607	4674	-	1068	1068	GENE	660.3	100	<input checked="" type="checkbox"/>
6 06		0	4756	5583	-	828	828	GENE	505.9	99.6	<input checked="" type="checkbox"/>
7 07		0	6056	6211	-	156	156	GENE	96.9	100	<input checked="" type="checkbox"/>
8 08		0	6284	6772	-	489	489	GENE	310.4	99.4	<input checked="" type="checkbox"/>

atgtcatcattaaacttttttctatttttctgttttcaactatctatttttactgtactgttaagcgggaatatcaataaagcgtgcacagaaagaaatatacatggaatcgaagtag
M S B L T P F I P C L Q L F I F T S C V S G I S I K R C T E E E N N T W E I E V Q

Annotations Blast Alignment(s) Needle Alignment

New Annotations ORFs & Reference Genes

Genome Map



Ref: SPPV_02 [1/1] complement(1232..2167)

1232 2167

Reset

Jump

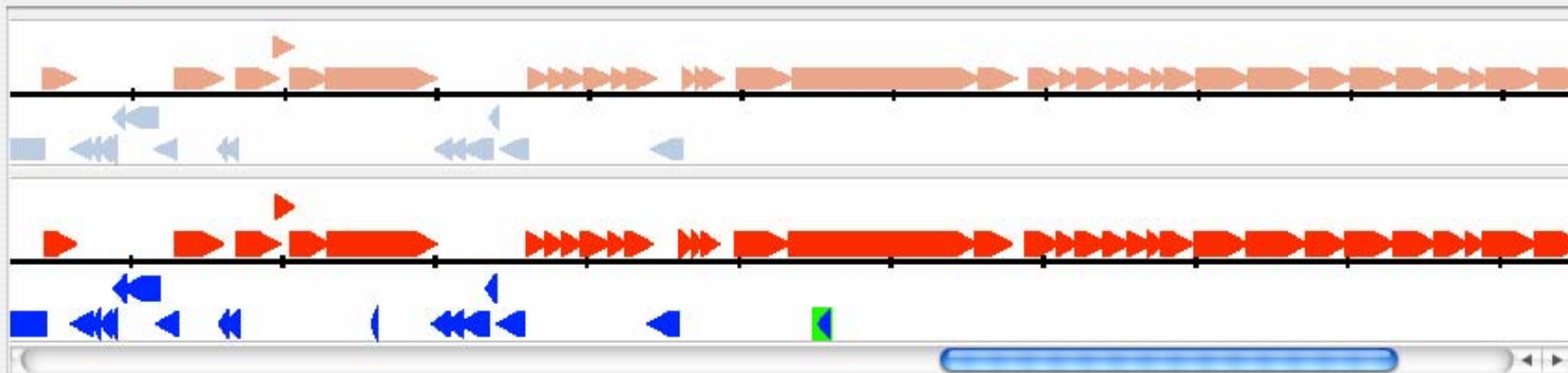
Annotate

VGO

BaseByBase

Save

Genome Map



Reset

Jump

Annotate

VGO

BaseByBase

Save

Annotation of bacterial genomes with GATU

2 strains

Chlamydia pneumoniae (TW183) annotated using strain AR39 as reference.
>90% nucleotide identity.

AR39 has 1112 annotated genes

1069 ORFs (96%) were accepted automatically by GATU.

Of the remaining 42 of 43 others were *Fragments*

13 new ORFs were added to TW183 TOTAL=1124 (1113 GenBank)

25' for GATU analysis

2 species

Thermoplasma volcanium (GSS1) annotated using *T. acidophilum* DSM-1728.

DSM-1728 has 1482 annotated genes

1268 ORFs (85%) were accepted automatically: 11 removed (duplicates)

80 ORFs (<60% similarity) accepted after manual review

1337 out of 1482 *T. acidophilum* genes transferred to target genome

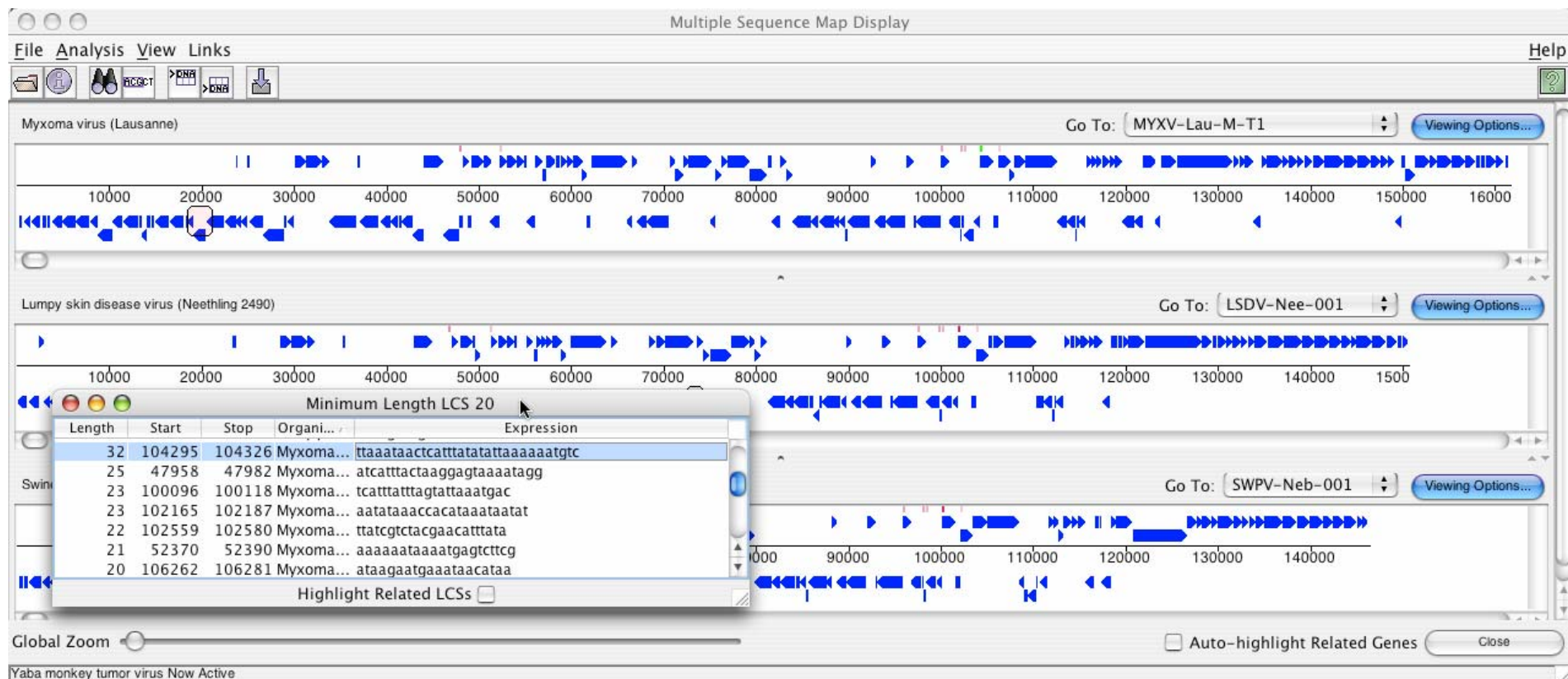
40' for GATU analysis

Longest Common Strings (LCS):

Finding common sequences in unaligned genomes.

40 nt region described to be conserved in set of poxviruses
- possibly promoter or origin of replication.

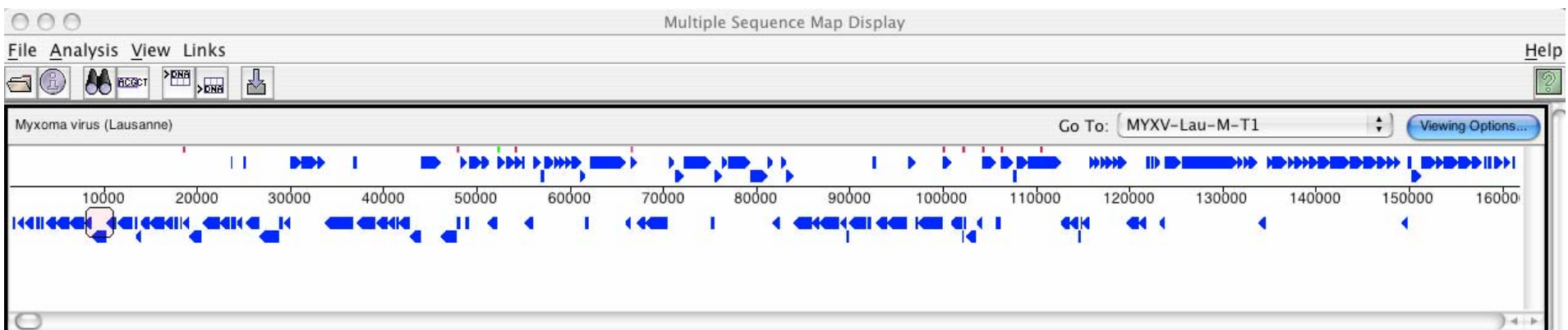
Are there others? - YES; most appear to be promoters.

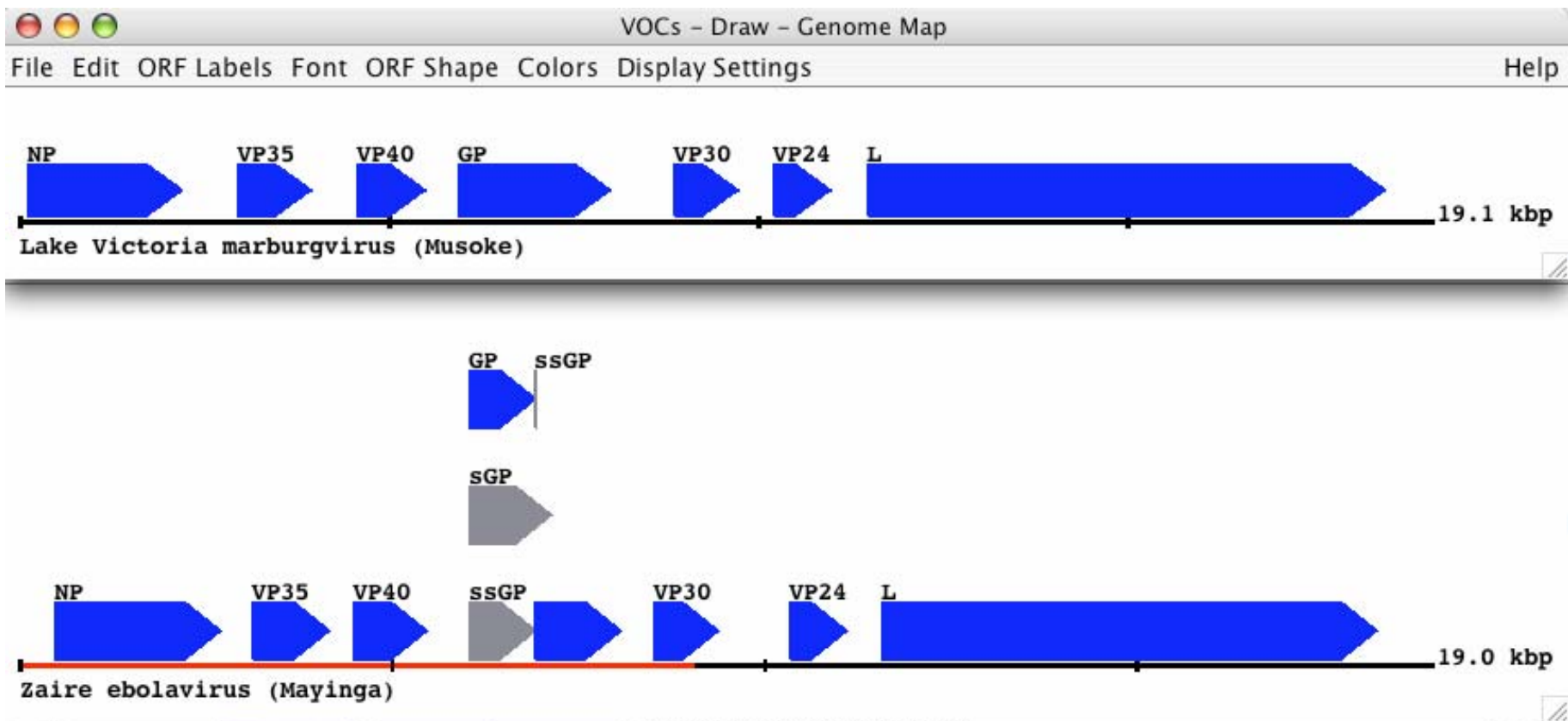


6 genomes: LCS>19 nt.

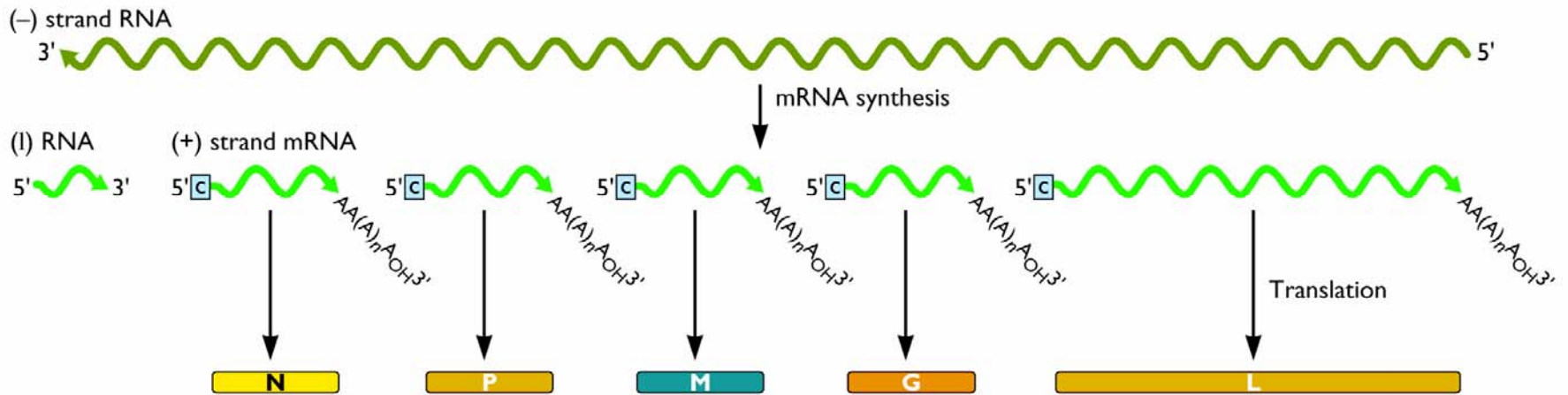
Annotation of nucleotide ambiguities in viral meta-genomes:

- 1) Multiple genomes of highly variable viruses.
if virus cloned, each gives single sequence
but need to represent consensus of quasi-species
- 2) Sequences of pooled genomes (e.g. viral RNA from liver).
each sequence read has ambiguities
need to compare different pools (virus from brain vs liver)





Filoviruses replicate like Rhabdoviruses:



Polyadenylation signal sequence between each gene:
AUACUUUUUUUU

Dot Plot Results

